Airline Customer Sentiment Analysis Report

1. Introduction

Sentiment Analysis is a prominent application of Natural Language Processing (NLP) that enables the classification of textual data based on emotions or opinions. It plays a vital role in understanding customer feedback, identifying issues, and improving services. In this project, we focus on analyzing customer reviews for airlines, categorizing them into positive, negative, and neutral sentiments, and identifying the underlying themes through topic modelling.

Additionally, we employ machine learning classification models to predict sentiments based on review data, thereby streamlining decision-making processes.

2. Problem Statement

The objective is to analyze airline reviews to:

- Classify reviews into positive, negative, or neutral sentiments.
- Identify actionable themes and patterns in customer feedback using topic modelling.
- Build and compare classification models to predict sentiments, selecting the best-performing model based on evaluation metrics such as F1-Score.

3. Data Preprocessing and Story Generation

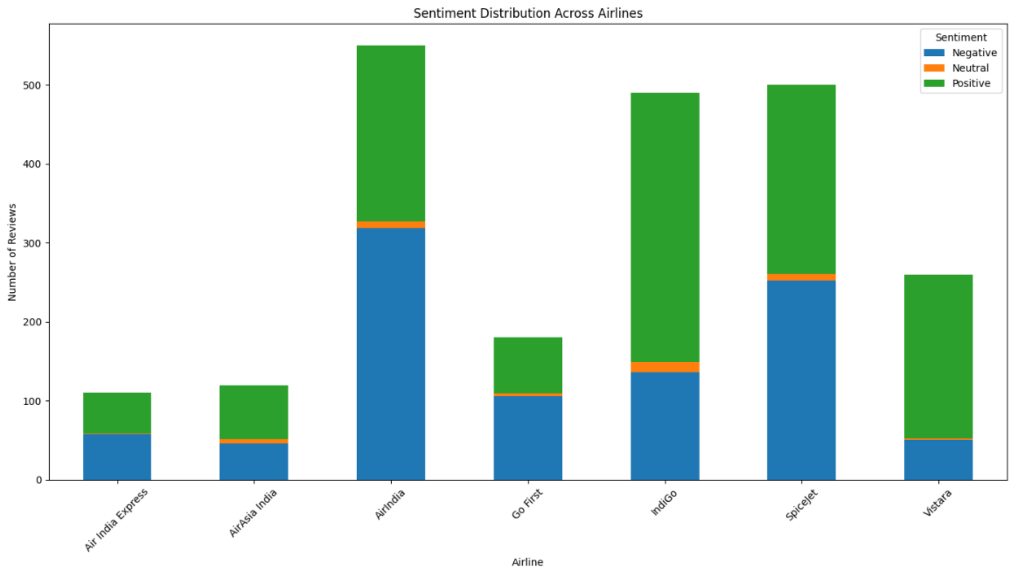
Preprocessing raw text data is essential to improve the quality of feature extraction. The following steps were performed:

- **Noise Removal**:
    - Removed special characters, punctuation, and numeric values.
    - Eliminated irrelevant terms like stopwords (e.g., "the", "is").
- **Lowercasing**:
    - Converted all text to lowercase to ensure consistency.
- **Tokenization and Lemmatization**:
    - Split reviews into tokens (words).
    - Reduced words to their base forms (e.g., "running" → "run") using lemmatization.
- **Custom Stopword Removal**:
    - Added airline-specific words (e.g., "flight", "airline") to the stopword list to reduce noise.

These steps ensured the creation of a high-quality feature space, essential for accurate modelling.

4.Exploratory Data Analysis (EDA) Observations

4.1 Sentiment Distribution:



- Positive sentiments dominated reviews for airlines like **Vistara** and **IndiGo**, indicating higher customer satisfaction.
- Negative sentiments were prevalent in reviews for **Air India** and **SpiceJet**, highlighting operational inefficiencies.
- Neutral sentiments were minimal across airlines.

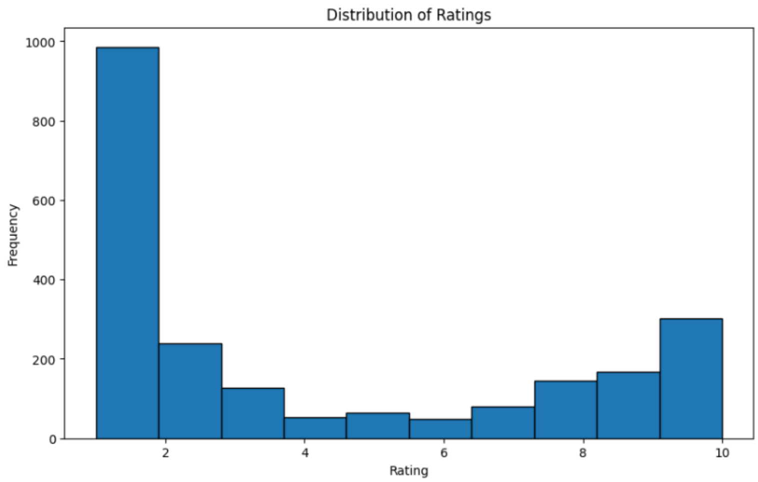
4.2 Frequent Words and Phrases:

- Positive reviews often included terms like **"friendly staff"**, **"on-time"**, and **"clean"**.
- Negative reviews emphasized phrases such as **"flight delayed"**, **"poor service"**, and **"lost baggage"**.

4.3 Word Clouds:

- Generated visualizations for frequent terms in positive and negative reviews to uncover patterns.

4.4 Distribution of Ratings

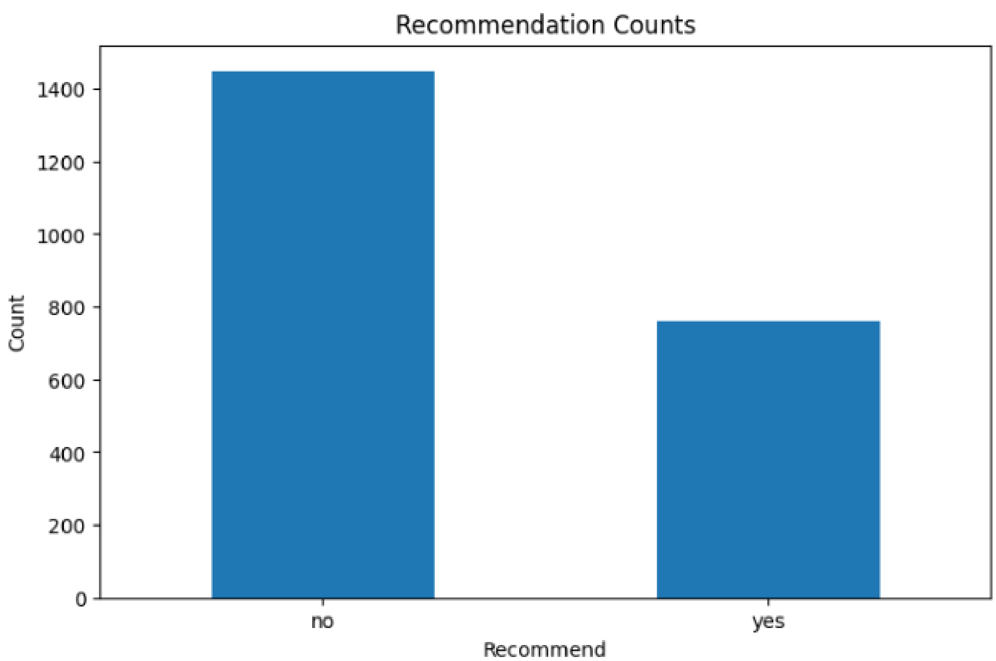


Observation:

The majority of ratings are concentrated at **1 and 2**, indicating a large proportion of highly dissatisfied customers.
There is a visible rise in ratings at **8 and 10**, showing a subset of highly satisfied customers.
Ratings in the middle range **(4–6)** are less frequent, suggesting polarized customer experiences.

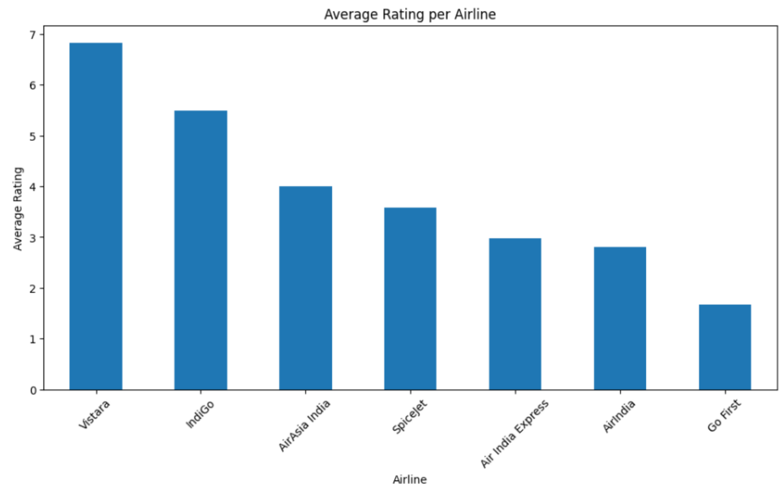
4.5 Recommendation Trends:

Airlines with lower ratings also have higher proportions of 'No' recommendations.



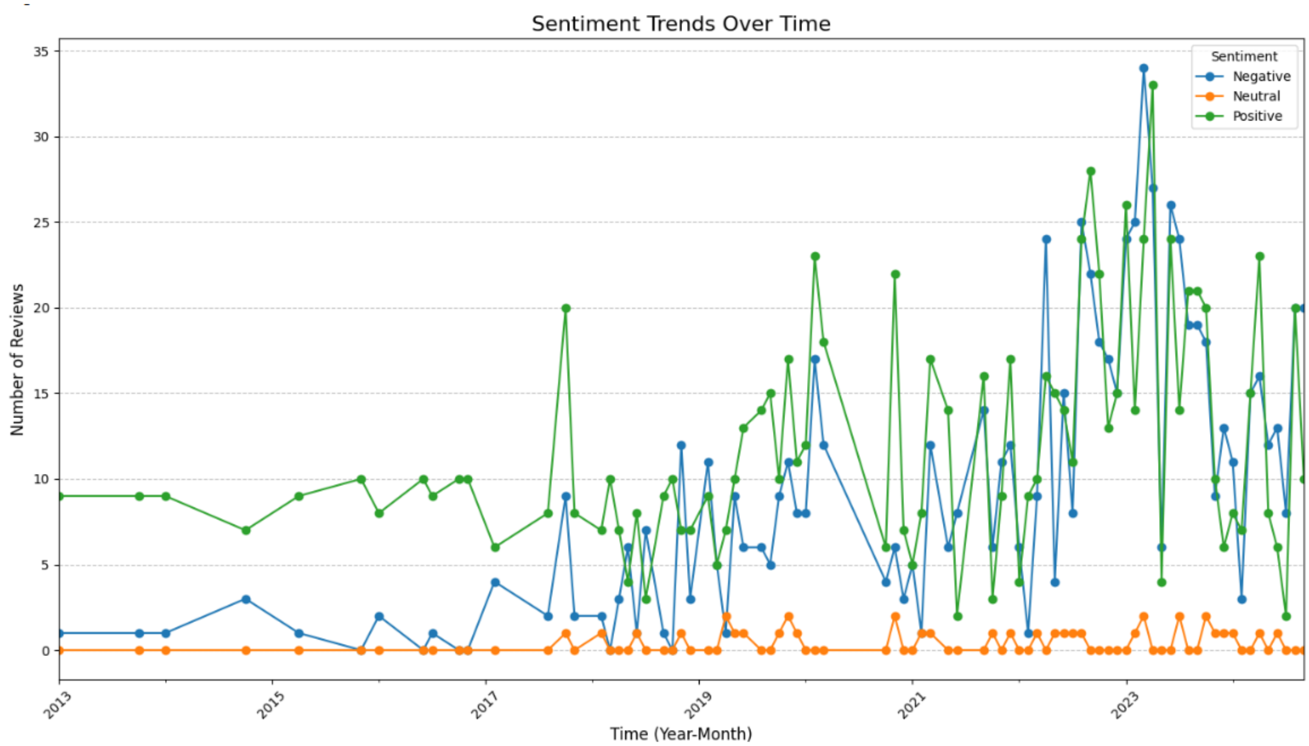
4.6 Average Rating Per Airline:

Airlines like Vistara and IndiGo consistently score higher in average ratings.Airlines like Air India and SpiceJet show lower average ratings.



5.Sentiment Analysis Observations

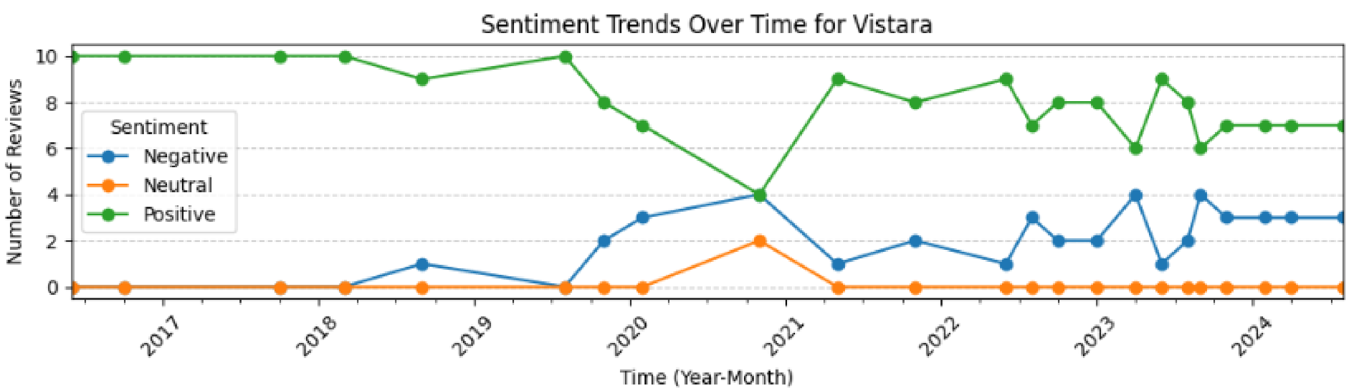
5.1 Inference for Sentiment Trends Over Time

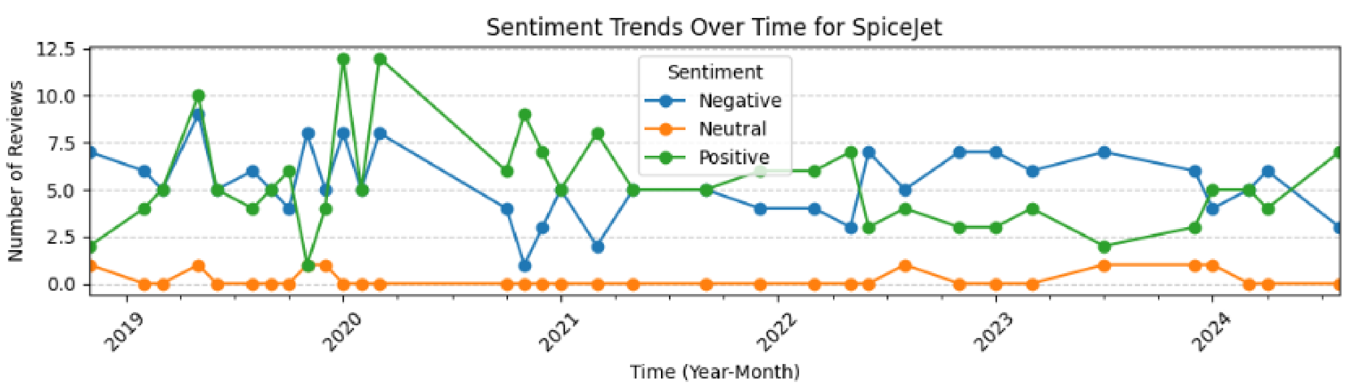


- **Positive Sentiment Dominance:** The green line representing positive sentiment consistently stays higher compared to negative and neutral sentiments. This indicates that, over time, customers generally had more positive experiences.

- **Fluctuations in Sentiment:** There are noticeable spikes and dips in positive sentiment, particularly around 2019 and 2023, which could correlate with specific events or improvements in services offered by airlines.
- **Negative sentiment (blue line)** also shows periodic increases, possibly linked to operational challenges, delays, or service failures during certain months.
- **Neutral Sentiment Stability:** Neutral sentiments (orange line) remain consistently low throughout the timeline, suggesting that most reviews lean toward definitive opinions (positive or negative).
- **Trends Around 2020–2021:** There is a dip in all sentiment types during this period, which could be attributed to the global reduction in air travel during the COVID-19 pandemic.
- **Increase in Engagement Post-2021:** A significant rise in both positive and negative sentiments after 2021 indicates a resurgence in air travel and increased customer feedback. Airlines may have introduced new services or faced challenges scaling operations during the recovery period.

5.2 **Sentiments Trend over time Airline wise**





- **Positive Sentiments**: Airlines like IndiGo and Vistara show the highest consistency in positive trends, indicating strong customer satisfaction.
- **Negative Sentiments:** AirIndia and SpiceJet exhibit significant spikes, reflecting potential operational or service-related challenges during specific periods.

- **Neutral Sentiments:** Across all airlines, neutral sentiments are consistently low, showing that most customers tend to share decisive opinions.

6.Topic Modeling Observations

Using Latent Dirichlet Allocation (LDA), we identified key topics for each airline:

- AirAsia India:
  - Baggage handling, ticketing issues, and staff behavior were recurring themes.
- Air India:
  - Delays, customer care inefficiencies, and flight cancellations dominated reviews.
- Vistara:
  - Premium services like food and seating received positive mentions, with occasional complaints about baggage.
- SpiceJet:
  - Frequent issues with cancellations, delays, and poor customer service were observed.

Incorporating bigrams and trigrams enhanced the contextual relevance of identified topics.

7. Model Performance and evaluation

We built and evaluated several machine learning models to classify sentiments:

- **Logistic Regression**:
  - Performed well on linearly separable data.
- **Decision Tree**:
  - Captured non-linear patterns but tended to overfit.
- **Random Forest**:
  - Combined multiple decision trees for high accuracy and robustness.
- **Support Vector Machine (SVM)**:
  - Performed well on high-dimensional data but required careful tuning.
- **Neural Network (MLP)**:
  - Captured complex relationships in data but required significant computational resources

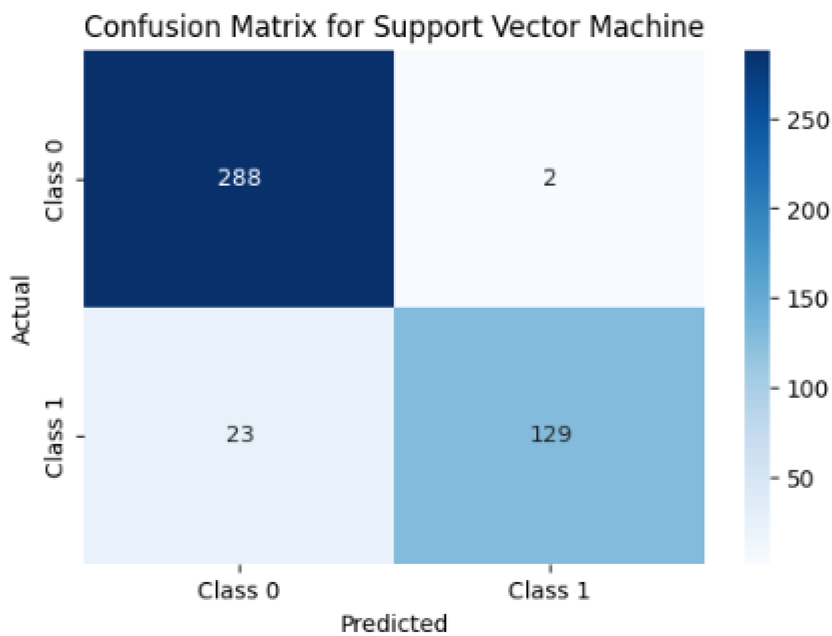
Performance Metrics:

On the basis of model evaluation Support Vector Machine showed the highest accuracy the outcomes of which are as follows

```
Classification Report for Support Vector Machine:
              precision    recall  f1-score   support

           0       0.93      0.99      0.96       290
           1       0.98      0.85      0.91       152

    accuracy                           0.94       442
   macro avg       0.96      0.92      0.94       442
weighted avg       0.95      0.94      0.94       442
```



Confusion Matrix Insights:
High true positive rate, indicating accurate predictions.
Some misclassifications in borderline cases.
Actionable Insight:
The model can reliably predict customer recommendations, helping airlines identify potential churn risks.

8. Business Insights: Actionable Insights Airline-Wise

A.AirAsia India:
Common Issues:
- Frequent complaints about baggage handling and delays.
- Staff behavior and ticketing issues often highlighted negatively.

Recommendations:
- Improve baggage management systems and train staff to handle disputes better.
- Simplify the ticketing process and offer compensation for delays to improve customer perception.

B. Air India:
Common Issues:
- Delays, poor customer service, and frequent cancellations dominate negative reviews.
- Positive mentions focus on business class and in-flight entertainment.

Recommendations:

- Invest in operational efficiency to minimize delays and cancellations.
- Enhance customer care processes with faster grievance redressal.
- Promote business class offerings through targeted marketing.

C. Air India Express:

Common Issues:

- Leg space and seating comfort often criticized.
- Persistent issues with delayed flights and baggage mishandling.

Recommendations:

- Reassess aircraft configuration to improve leg space.
- Implement stricter punctuality measures and streamline baggage handling processes.

D. Go First:

Common Issues:

- Complaints about cancellations, extra charges, and rescheduling.
- Poor reliability and unprofessional customer care.

Recommendations:

- Clearly communicate policies regarding charges and cancellations.
- Introduce proactive notifications for rescheduling to reduce customer frustration.
- Strengthen customer service training programs to improve professionalism.

E. IndiGo:

Common Issues:

- Criticism of ground staff and limited support for senior citizens.
- Positive reviews emphasize affordability and cleanliness.

Recommendations:

- Develop senior citizen-friendly services, such as priority boarding and assistance.
- Address ground staff inefficiencies with targeted training programs.
- Highlight affordability and cleanliness in marketing campaigns to attract budget travelers.

F. SpiceJet:

Common Issues:

- Significant dissatisfaction with delays and schedule changes.
- Concerns about senior citizen assistance and value for money.

Recommendations:

- Enhance scheduling accuracy and establish a clear compensation policy for delays.
- Focus on providing inclusive services, especially for senior citizens.
- Address the perception of low value for money by improving basic services like seating and in-flight amenities.

G. Vistara:

Common Issues:

- Minor complaints about baggage handling and occasional delays.
- Strong positive feedback on premium services, including food and seating.

Recommendations:

- Maintain premium service quality and focus on consistent delivery to retain customer trust.
- Optimize baggage handling processes to reduce occasional issues.
- Highlight premium offerings in marketing campaigns to reinforce the brand's premium positioning.

9. Conclusion

This analysis provides actionable insights into customer sentiment, recurring themes, and predictive analytics for Indian domestic airlines. By implementing the recommendations derived from sentiment analysis, topic modeling, and logistic regression, airlines can improve operational efficiency, enhance customer satisfaction, and reduce churn rates effectively.